Scalability of the VSO

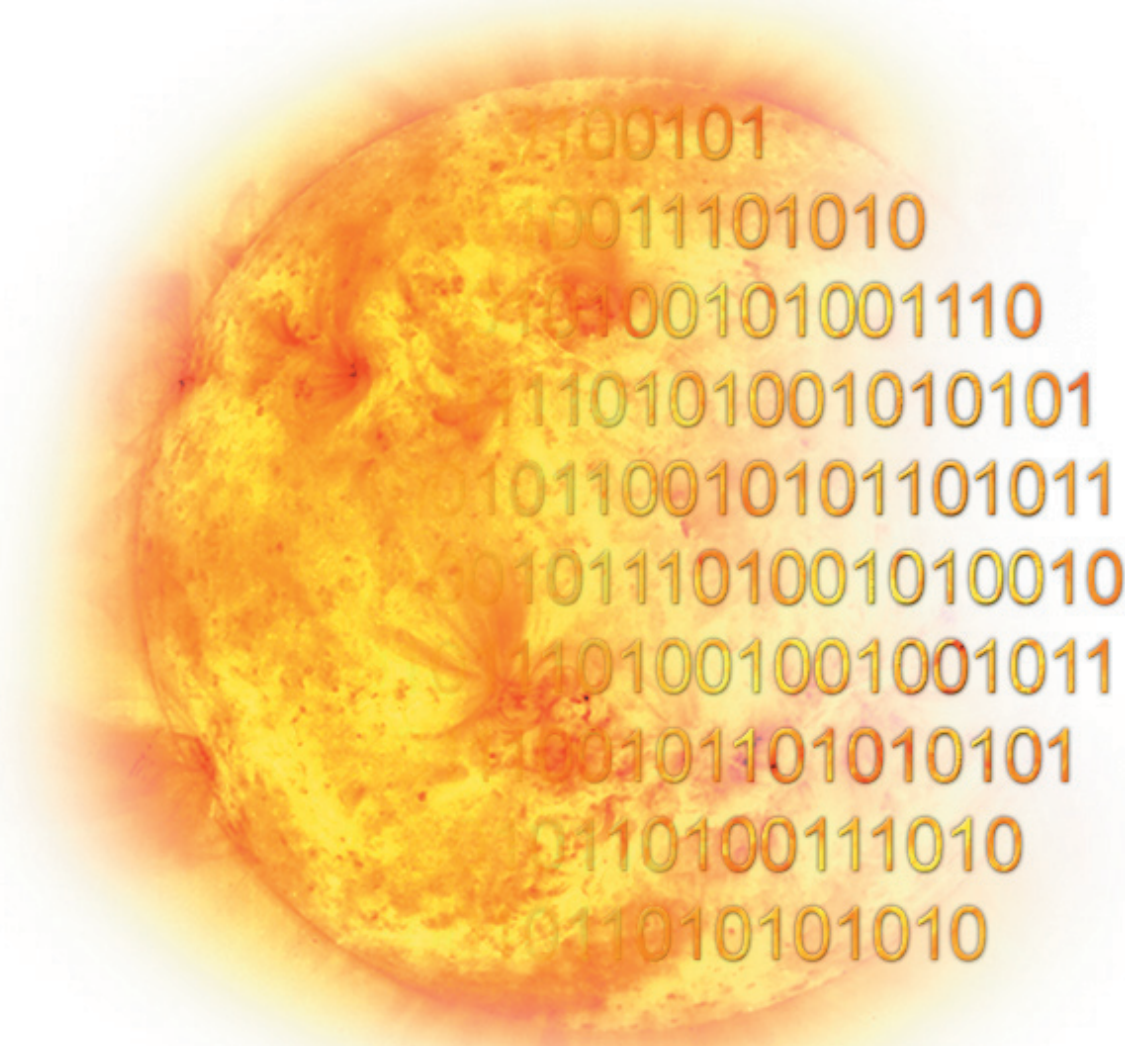J.A. Hourclé (GSI/NASA GSFC) A.R. Davey (SwRI) I. Suarez-Sola (NSO) K.Q. Tian (Stanford U.) K. Yoshimura (MSU-Bozeman)

Virtual Solar Observatory

Abstract:

The successful development of the VSO prototype raises fundamental system architecture issues in dealing with the scalability and fault tolerance of distributed systems. The four sites participating in development (Stanford, MSU, NSO and SDAC) are only the starting point of VSO.

The system's overall architecture is being influenced by interaction issues between different data repositories that do not share a common set of capabilities. As new data providers or new search capabilities are added, the time needed to return useful results should not grow linearly with them. We take care to identify and isolate slow or unresponsive data repositories, to keep them from adversely affecting the overall quality of VSO searches.

This extra complexity needs to remain unobtrusive to the solar researches, so that they can focus on their research, and not the tools that they are using.

Scalability:

The modular design used by VSO allows the addition of new features or functionality without impacting unrelated components. The design provides for horizontal scalability by allowing each VSO component, except for the individual Data Providers, to be implemented on more than one system to distribute the workload.

If desired, the individual instances of components may also be vertically scaled, by upgrading the systems that are hosting them. Depending on the performance required, each individual module may be hosted on a separate server, or a single server may contain multiple modules across multiple logical containers.

Modularity:

Each individual component of VSO performs simple tasks which provides the overall complex functionality to an end user.

Central Server:
Handles distribution and collection of information between VSO Instances.
Registry:
Provides directory information about the different Data Repositories.
Data Query Interface:
Translates VSO queries to a format that a Data Repository can understand.
Data Repository:
An online service containing solar data.
User Interface:
Handles the presentation to the user
User Query Interface:
Translates queries to the format used internally by VSO
VSO Core:
Handles the multiplexing of the queries to multiple data providers.
Logical Containers:

Access Method:
The components that translate a User's intent into a VSO query.
Data Provider:
An organization that hosts one or more Data Repositories.
VSO Instance:
An application that proxies requests between Users and Data Providers

Monitoring:

To ensure that users are given the most comprehensive search possible, in a reasonable amount of time, it is necessary to monitor the availability of individual components. If a component seems to be performing outside the range of accepted behaviour, one of the VSO Central Servers will determine its availability, and should it be unresponsive, will watch for it to return to an acceptable state.

Fault Tolerance:

By breaking the tasks down into smaller components, we create self contained items which are horizontally scalable. As the VSO design does not require the maintenance of session state on a component, we can use a failover design to provide high availability of VSO components without the overhead of a more complex clustered environment.

By maintaining information about components in the VSO Registry, we can easily add information about new Data Repositories, or temporarily skip over a Data Repository that might be down for maintenance or otherwise unavailable or unresponsive.

Tuning:

The simple design of the VSO components allows for each component to be individually tuned to receive the best performance characteristics possible. The simplicity of the Registry design allows us to use vague descriptions to quickly get information available online, or to use more specific descriptions to reduce the number of false positives that might otherwise occur.

The centralized logging capabilities allow the maintainers to review the logs across multiple VSO instances. From these logs, they can determine when and where false positives are occurring, and adjust the Registry entries to reduce unnecessary queries that might otherwise be sent to a Data Provider. Through proper tuning, we can reduce the negative impact that may otherwise be caused to a Data Provider by being a part of a Virtual Observatory.